Supplementary Notes

NMR and EC-derived residue contact information can be highly complementary. NMR data obtained using perdeuterated, selectively reprotonated protein samples, primarily from NOESY experiments[1-7], is incomplete and often ambiguous in its assignment to specific ^{1}H-^{1}H interactions. None the less, all (or most) of these NOE data should be consistent with the 3D structure model(s). EC-based contacts can complement this spectroscopic information to provide more contact information, and more accurate models, but potentially include predicted interactions that are false positives (FPs) for the subject protein. The requirement that the overall structure be consistent with all of the experimental NMR data, however, provides "hard" constraints on the interpretation of ECs, allowing identification and removal of such FP residue pair contacts.

In certain favorable cases, extensive backbone and sidechain resonance assignments have been determined for larger (24 – 65 kDa) proteins using primarily ^{13}C,^{15}N-edited NOESY, HNCA, and ^{13}C total correlation spectroscopy, allowing structure determination without the requirement of perdeuteration[8]. It is anticipated that such strategies would also be enhanced by incorporation of evolutionary information, together with such experimental NMR data, in assigning NOESY cross peaks and in generating accurate protein structures. It is often the case in protocol development that cutoffs and numbers of iterations are assessed and adopted even though they may not represent the most optimum set of parameters. This first description the hybrid EC-NMR method will have a tremendous impact and influence on the field of protein NMR structure determination, using a hybrid combination of evolutionary and NMR data, and on structural biology in general.

In all of the calculations we carried out, the EC-NMR protocol increased the number of ECs that could be used reliably in structure calculations and reduced the FP rates (**Fig. 2a**, Supplementary Figure 7 and Supplementary Table 4). For example, for MBP using RDC data available for a single hydrodynamic alignment, the number of FPs was reduced from 127 to 82 and the Precision of the ECs (relative to the reference X-ray crystal structure) was increased from 0.66 to 0.73 (MALE_ECOLI in Supplementary Table 4B). The remaining false-positive EC contacts remaining at the end of the EC-NMR protocol are nearby in the 3D structures to true-positive inter-residue contacts (as can be see visually in the final contact shown in the bottom panel of **Fig. 2**), and have little effect on the accuracy of the final structure. Similarly for the other EC-NMR structures that were calculated, all of the FPs that are significantly inconsistent with the reference structures were eliminated (**Fig. 2**; Supplementary Fig. 7, and Supplementary Table 4B). Hence the EC-NMR protocol is very effective in removing structurally-inaccurate FPs from the EC lists.

A prerequisite for the EC-NMR method is extensive, diverse sequence data needed to obtain accurate co-evolutionary couplings between residue pairs[9-11]. The sensitivity analysis of P74712 using various depth MSAs indicated that 5*L non-redundant sequences (N_{eff}) are required for accuracy < 3.5 Å RMSD. For a target protein that is 200 residues long, this typically requires on the order of 5,000 sequences, before removal of redundancy, in an initial multiple sequence alignment[11-15]. These estimates of sequence-depth requirements clearly depend on the quality of the sparse NMR data and the availability of RDC data. With current sequence databases, approximately 50% of known protein domain families could be addressed using the EC-NMR method today[12]. Although these domain families cover only a fraction of all proteins, proteins become new domain families, or merge with others, as increasing sequence information becomes available. The size of the sequence databases is currently doubling every two years: from 2013 to 2014 the number of proteins increased from 40 million to 90 million. One can thus expect that each year increasing numbers of proteins will be amenable to the EC-NMR approach.

Supplementary Table 1. Statistics for ECs used in EC-NMR analysis.

Protein ID	Length covered[1]	Fraction of sequence covered by ECs (%)	Number of non-redundant sequences in MSA	Number of sequences per covered residue[2]
A9CJD6_AGRT5	63 (80)	79	10,962	174
Q6D6V0_ERWCT	63 (75)	84	4,410	70
Q9ZV63_ARATH	73 (91)	80	4,964	68
Q1LD49_RALME	131(134)	97	2,620	20
YIAD_ECOLI	132 (142)	93	10,296	78
RASH_HUMAN	171 (189)	90	6,669	39
P74712_SYNY3	156 (194)	79	45,708	293
MALE_ECOLI	388 (396)	98	12,416	32

[1] Number of residues covered with ECs (full length of protein used in query).

[2] Number of sequences after filtering and reducing for redundancy divided by number of residues covered.

Supplementary Table 2. Experimental protein NMR data sets.

Protein Uniprot ID	PDB DOI PDB_ID BMRB_ID	No. Restraints / Residue Total	Long Range	No. N-H RDCs Alignment 1	Alignment 2	PDB Author List
A9CJD6_AGRTT	10.2210/pdb2k2p/pdb 2K2P 15721	16.53	3.86	N/A	N/A	Lemak, A., Gutmanas, A., Yee, A., Semesi, A., Arrowsmith, C.H.
Q6D6V0_ERWCT	10.2210/pdb2k5n/pdb 2K5N 15843	14.55	4.86	59	59	Mills, J.L., Eletsky, A., Zhang, Q., Lee, D., Jiang, M., Ciccosanti, C., Xiao, R., Lui, J., Everett, J.K., Swapna, G.V.T., Acton, T.B., Rost, B., Montelione, G.T., Szyperski, T
Q9ZV63_ARATH	10.2210/pdb2kan/pdb 2KAN 16029	20.30	6.74	68	48	Eletsky, A., Mills, J.L., Sukumaran, D., Hua, J., Shastry, R., Jiang, M., Ciccosanti, C., Xiao, R., Liu, J., Everret, J.K., Swapna, G.V.T., Acton, T.B., Rost, B., Montelione, G.T., Szyperski, T.
Q1LD49_RALME	10.2210/pdb2lcg/pdb 2LCG 17611	17.51	6.79	122	N/A	Ramelot, T.A., Yang, Y., Cort, J.R., Wang, D., Ciccosanti, C., Janjua, H., Rajesh, N., Acton, T.B., Xiao, R., Everett, J.K., Montelione, G.T., Kennedy, M.A.
YIAD_ECOLI	10.2210/pdb2k1s/pdb 2K1S 16293	7.47	3.44	125	116	Ramelot, T.A., Zhao, L., Hamilton, K., Maglaqui, M., Xiao, R., Liu, J., Baran, M.C., Swapna, G., Acton, T.B., Rost, B., Montelione, G.T., Kennedy,
RASH_HUMAN	10.2210/pdb2lcf/pdb 2LCF 17610	18.64	6.29	N/A	N/A	Araki, M., Shima, F., Yoshikawa, Y., Muraoka, S., Ijiri, Y., Nagahara, Y., Shirono, T., Kataoka, T., Tamura, A.
P74712_SYNY3	10.2210/pdb2kw5/pdb 2KW5 16806	5.14	2.47	108	108	Lange, O.F., Rossi, P., Sgourakis, N.G., Song, Y., Lee, H.W., Aramini, J.M., Ertekin, A., Xiao, R., Acton, T.B., Baker, D., Montelione, G.T.
MALE_ECOLI[1]	10.2210/pdb1ezp/pdb 1EZP 4986	5.25	2.16	280	N/A	Mueller, G.A., Choy, W.Y., Yang, D., Forman-Kay J.D., Venters, R.A., Kay, L.E.
	10.2210/pdb2mv0/pdb 2MV0 25237	5.00	1.98	N/A	N/A	Rossi, P., Lange, O.F., Sgourakis, N.G., Song, Y., Lee, H., Aramini, J.M., Ertekin, A., Xiao, R., Acto T.B., Baker, D., Montelione, G.T.

1. For EC-NMR calculations on MBP (MALE_ECOLI) the NOESY peak list data are from PDB_ID 2MV0, and the ^{15}N-^{1}H RDC data are from PDB_ID 1EZP.

Supplementary Table 3. Comparison of NOE-based restraint densities in EC-NMR and NMR reference structures.

Protein Uniprot ID	Number of RDC alignments	No. Restraints / Residue for EC-NMR structure		Reference NMR PDB_ID	No. Restraints / Residue for reference NMR structure	
		Total	Long_Range		Total	Long_Range
A9CJD6_AGRTT5	2	0.56	0.09			
	0	0.58	0.09	2K2P	16.53	3.86
Q6D6V0_ERWCT	2	1.20	0.15	2K5N	14.55	4.86
Q9ZV63_ARATH	2	0.92	0.10	2KAN	20.30	6.74
Q1LD49_RALME	2	1.04	0.15			
	1	1.04	0.15	2LCG	17.51	6.79
YIAD_ECOLI	2	1.11	0.12	2K1S	7.47	3.44
RASH_HUMAN	2	1.78	0.16			
	0	1.86	0.17	2LCF	18.64	6.29
P74712_SYNY3	2	5.04	2.03	2KW5	5.14	2.47
MALE_ECOLI	2	5.08	1.92			
	1	5.09	1.91	2MV0	5.00	1.98

Supplementary Table 4. Statistics assessing performance of EC-NMR protocol in identifying false positive and false negative Residue Pair Contacts[a] (RPCs).

A. Recall (R), Precision (F), and Performance (F) Scores for initial EC and final RPC lists

Protein Uniprot ID	No. of RDC hydrodynamic alignments	Initial EC List				Final RPC List[a]			
		N	R	P	F	N	R	P	F
Smaller Proteins (< 15 kDa)									
A9CJD6_AGRTT5	2	65	0.34	0.75	0.47	74	0.42	0.81	0.55
	0					77	0.41	0.75	0.53
Q6D6V0_ERWCT	2	75	0.24	0.60	0.34	102	0.44	0.81	0.57
Q9ZV63_ARATH	2	85	0.30	0.72	0.42	110	0.43	0.86	0.60
Q1LD49_RALME	2	139	0.27	0.78	0.40	180	0.39	0.86	0.53
	1					179	0.38	0.85	0.53
YIAD_ECOLI	2	150	0.24	0.57	0.34	180	0.35	0.69	0.48
Larger Proteins (> 15 kDa)									
RASH_HUMAN	2	167	0.27	0.63	0.38	221	0.42	0.73	0.53
	0					227	0.43	0.72	0.53
P74712_SYNY3	2	198	0.20	0.62	0.30	413	0.47	0.69	0.56
MALE_ECOLI	2	371	0.18	0.66	0.28	761	0.41	0.74	0.53
	1					739	0.39	0.73	0.51

[a]RPC - Residue Pair Contacts. These are contacts between pairs of residues deduced by combined analysis of the initial EC list, NMR data, and intermediate structural data.

B. False Positives (FPs) in initial EC and final EC lists

Protein Uniprot ID	No. of RDC hydrodynamic alignments	Initial EC List			Final EC List	
		N	FP	P	FP	P
Smaller Proteins (< 15 kDa)						
A9CJD6_AGRTT5	2	65	16	0.75	12	0.81
	0				16	0.75
Q6D6V0_ERWCT	2	75	30	0.60	19	0.81
Q9ZV63_ARATH	2	85	24	0.72	15	0.86
Q1LD49_RALME	2	139	30	0.78	22	0.86
	1				23	0.85
YIAD_ECOLI	2	150	64	0.57	54	0.69
Larger Proteins (> 15 kDa)						
RASH_HUMAN	2	167	62	0.63	48	0.73
	0				55	0.72
P74712_SYNY3	2	198	76	0.62	50	0.69
MALE_ECOLI	2	371	127	0.66	72	0.74
	1				82	0.73

Supplementary Table 5. Assessment of the accuracy of well-defined, buried side chain χ_1 dihedral angles.

Protein NMR Data Set	Reference NMR Structure	Number of buried, well-defined sidechains[1]	χ_1 rotamer agreement (%)
A9CJD6_AGRT5	2K2P	13	80
A9CJD6_AGRT5*	2K2P	8	84
Q6D6V0_ERWCT	2K5N	5	99
Q9ZV63_ARATH	2KAN	17	79
Q1LD49_RALME	2LCG	20	85
Q1LD49_RALME**	2LCG	22	86
YIAD_ECOLII	2K1S	26	87

[1]Side chains that are buried (SASA < 40 Å^2 in the X-ray structure) and well-defined (χ1 angle S.D. < 30 degrees in the NMR ensemble).

* EC-NMR structure calculated based on RDCs computed for two alignments.

** EC-NMR structure calculated based on experimental RDCs for 1 alignment and computed RDCs for a second alignment.

p21 H-Ras

NMR Structure	Number of buried, well-defined sidechains	χ_1 rotamer agreement (%)	Number of common buried, well-defined sidechains[1]	χ_1 rotamer agreement (%)	RMSD to X-ray crystal structure[2] (Å)
2LCF	43	87	26	86	1.5
EC-NMR	15	85	26	73	2.6
EC-NMR**	40	87	26	92	1.6

[1]Side chains that are buried (SASA < 40 Å^2 in the X-ray structure) and well-defined (χ1 angle S.D. < 30 degrees in the NMR ensemble).

[2]The reference X-ray crystal structure is PDB ID 5P21.

**EC-NMR structure calculated based on RDCs computed for two alignments.

P74712

NMR Structure	Number of buried, well-defined sidechains	χ_1 rotamer agreement (%)	Number of common buried, well-defined sidechains[1]	χ_1 rotamer agreement (%)	RMSD to X-ray crystal structure[2] (Å)
2KW5	52	71	23	65	2.4
EC-NMR	28	81	23	77	2.1

[1]Side chains that are buried (SASA < 40 Å^2 in the X-ray structure) and well-defined (χ_1 angle S.D. < 30 degrees in the NMR ensemble).

[2]The reference X-ray crystal structure is PDB ID 3MER.

Maltose Binding Protein

NMR Structure	Number of buried, well-defined, side chains[1]	χ_1 rotamer agreement (%)	Number of common buried, well-defined sidechains[1]	χ_1 rotamer agreement (%)	RMSD to X-ray crystal structure[2] (Å)
2D21	105	76	15	57	5.2
1EZP	33	26	15	23	3.2
2MV0	80	75	15	60	4.7
EC-NMR	78	65	15	57	2.8
EC-NMR**	59	80	15	60	2.1

[1]Side chains that are buried (SASA < 40 Å^2 in the X-ray structure) and well-defined (χ_1 angle S.D. < 30 degrees in the NMR ensemble).

[2]The reference X-ray crystal structure is PDB ID 1DMB.

**EC-NMR structure calculated based on experimental RDCs for 1 alignment and computed RDCs for a second alignment.

Supplementary Table 6. Sensitivity analysis of the EC NMR method using reduced protein sequence data for generating ECs for protein P74712.

Effective number of sequences (N_{eff})[1]	Percentage of available sequences used in alignment (%)	N_{eff} / L	Total number of NOE-based restraints	Number of long range [| i - j | ≥ 5] NOE – based restraints	Number of NOE restraints per restrained residue	Distance violations / structure			RMS of distance violation / restraint (Å)	Maximum distance violation (Å)	RMSD to the X-ray Structure bb / heavy atom	Precision of top L ECs (5 Å cutoff)	Number of high confidence ECs	Precision of high confidence ECs (5 Å cutoff)
						0.1-0.2 Å	0.2-0.5 Å	> 0.5 Å						
44172	100	227.69	978	397	5.8	7.4	5.2	3.05	0.05	1.59	2.2/3.6	0.50	89	0.67
24885	50	128.27	971	418	5.7	6	3.35	0.7	0.03	0.87	2.6 / 3.5	0.49	84	0.69
13641	25	70.32	974	397	5.8	5.05	4	1.85	0.07	1.69	2.6 / 3.6	0.50	80	0.71
11247	20	57.98	963	399	5.7	6.6	8.55	4	0.05	1.57	3.0 /4.0	0.47	57	0.70
8630	15	44.49	959	404	5.7	4.75	3.9	1.6	0.04	1.50	2.3 / 3.2	0.49	60	0.70
5994	10	30.90	957	372	5.7	6.1	6.5	2.25	0.04	1.01	3.0 / 4.1	0.46	57	0.68
3140	5	16.19	990	413	5.9	4.2	2.05	1.4	0.03	1.15	3.2 / 4.1	0.40	40	0.75
2547	4	13.13	996	403	5.9	7.2	4.7	2.3	0.05	1.25	3.5 / 4.2	0.39	43	0.72
1924	3	9.92	936	353	5.5	6.95	4.45	2.45	0.05	1.17	3.6 / 4.4	0.36	38	0.76
1307.	2	6.74	971	376	5.7	5.1	4.15	2.55	0.04	1.18	3.0 / 3.8	0.31	26	0.77
662	1	3.41	941	375	5.6	6.95	4.8	1.9	0.04	1.58	4.3 / 5.0	0.25	14	0.64
532	0.8	2.74	961	387	5.7	7.45	5.5	1.8	0.04	0.93	4.6 / 5.3	0.25	11	0.82
402	0.6	2.07	944	358	5.6	7.95	7.95	3.3	0.10	2.51	15.4 / 16.4	0.22	9	0.89
265	0.4	1.37	952	368	5.7	7.65	7.6	1.55	0.05	1.30	4.6 / 5.4	0.18	5	1.00
136	0.2	0.70	940	360	5.6	6.15	4.75	1.8	0.05	1.01	15.1 / 15.9	0.13	4	1.00
68	0.1	0.35	890	310	5.3	7.75	6.55	4.15	0.08	1.20	14.8 / 15.6	0.08	1	1.00
35	0.05	0.18	924	332	5.5	6.15	4.65	1.95	0.06	1.24	7.2 / 7.9	0.06	0	–
14	0.02	0.07	894	309	5.3	7.7	4.7	1.7	0.05	1.06	14.1 / 16.7	0.06	1	0.00
8	0.01	0.04	917	333	5.5	10.4	8.05	4.4	0.21	6.68	12.9 / 13.3	0.02	0	–
0	0	0	942	383	5.6	35.95	25.2	9.5	0.09	1.66	4.6 / 5.2	–	–	–

1. In main text Table 1 the N_{eff} for P74712 is 45,708, but in this sensitivity analysis it is 44,172. The sensitivity analysis was done using a more recent version of the EVfold pipeline with improved methods for defining non-redundant sequences for constructing multiple sequence alignments.

Supplementary Data 1. Python code for computing calculate EC reliability scores and the number of high-confidence ECs

```
#!/usr/bin/env python
"""
Calculate reliability score Q(i,j) for all EC scores in
an EVcouplings *_ECs.txt file. Also sorts the file
from largest to smallest EC scores (output to stdout).

Also outputs the number of long-range pairs with
Q(i,j) > 2 to stderr.

Date: 14.03.2015
"""

from sys import argv, stderr, exit

HIGH_CONFIDENCE_THRESHOLD = 2.0
SEQUENCE_SEPARATION = 5

def calculate_reliability_score(ec_file):
    """
    Calculate reliability scores for each EC
    in ec_file and print to stdout.

    Then counts number of EC pairs above score threshold.

    ec_file (str): *_ECs.txt file from EVcouplings
    """
    with open(ec_file) as f:
        ecs = [line.rstrip().split() for line in f]

    # sort the EC list by final column (EC score)
    ecs.sort(key=lambda x: float(x[-1]), reverse=True)

    # find the most negative EC score to estimate
    # level of background coupling
    min_abs = abs(float(ecs[-1][-1]))

    # calculate reliability score for each EC and output
    pairs_exceeding_threshold = 0
    sequence_indices = set()

    print " ".join(["i", "aa_i", "j", "aa_j", "ec_score", "ec_reliability"])
```

```
    for i, aa_i, j, aa_j, _, ec_score in ecs:
        print " ".join([i, aa_i, j, aa_j, ec_score]),
        print "{:.6f}".format(float(ec_score) / min_abs)

        sequence_indices.add(i)
        sequence_indices.add(j)
        if (float(ec_score) / min_abs > HIGH_CONFIDENCE_THRESHOLD and
                abs(int(i) - int(j)) > SEQUENCE_SEPARATION):
            pairs_exceeding_threshold += 1

    # output overall statistics to stderr
    indices_int = map(int, sequence_indices)
    domain_length = max(indices_int) - min(indices_int) + 1
    print >> stderr, "# Long-range EC pairs with Q_raw > 2:",
    print >> stderr, pairs_exceeding_threshold
    print >> stderr, "Domain length:", domain_length

if __name__ == "__main__":
    if len(argv) != 2:
        print >> stderr, "usage: {} <EC score file>".format(argv[0])
        exit(-1)

    ec_file = argv[1]
    calculate_reliability_score(ec_file)
```

Supplementary Data 2. *ASDP* commands for EC-NMR.

The *ASDP* commands used to run EC-NMR calculations are:

```
asdp  –c control-file   # Defines control file name required
      –o outputDir      # defines output file name
      –m                # defines that it is a sparse NMR data set (not fully protonated)
      –k 5.0            # calibration constant for converting NOE intensities to distances
      [–i]              # required only for perdeuterated sample,
                        #  instructs program to apply isotope shift corrections to
                        #  Ca and Cb chemical shift values.
```

References for Supplementary Materials

1 Lange, O. F. *et al.* Determination of solution structures of proteins up to 40 kDa using CS-Rosetta with sparse NMR data from deuterated samples. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 10873-10878, (2012).

2 Raman, S. *et al.* NMR structure determination for larger proteins using backbone-only data. *Science* **327**, 1014-1018, (2010).

3 Zheng, D. *et al.* Automated protein fold determination using a minimal NMR constraint strategy. *Protein science : a publication of the Protein Society* **12**, 1232-1246, (2003).

4 Hiller, S., Ibraghimov, I., Wagner, G. & Orekhov, V. Y. Coupled decomposition of four-dimensional NOESY spectra. *Journal of the American Chemical Society* **131**, 12970-12978, (2009).

5 Wang, X., Weldeghiorghis, T., Zhang, G., Imperiali, B. & Prestegard, J. H. Solution structure of Alg13: the sugar donor subunit of a yeast N-acetylglucosamine transferase. *Structure* **16**, 965-975, (2008).

6 Frueh, D. P. *et al.* Dynamic thiolation-thioesterase structure of a non-ribosomal peptide synthetase. *Nature* **454**, 903-906, (2008).

7 Sgourakis, N. G. *et al.* The Structure of Mouse Cytomegalovirus m04 Protein Obtained from Sparse NMR Data Reveals a Conserved Fold of the m02-m06 Viral Immune Modulator Family. *Structure* **22**, 1263-1273, (2014).

8 Xu, Y., Zheng, Y., Fan, J. S. & Yang, D. A new strategy for structure determination of large proteins in solution without deuteration. *Nature methods* **3**, 931-937, (2006).

9 Marks, D. S. *et al.* Protein 3D structure computed from evolutionary sequence variation. *PloS one* **6**, e28766, (2011).

10 Hopf, T. A. *et al.* Three-dimensional structures of membrane proteins from genomic sequencing. *Cell* **149**, 1607-1621, (2012).

11 Kamisetty, H., Ovchinnikov, S. & Baker, D. Assessing the utility of coevolution-based residue-residue contact predictions in a sequence- and structure-rich era. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 15674-15679, (2013).

12 Marks, D. S., Hopf, T. A. & Sander, C. Protein structure prediction from sequence variation. *Nature biotechnology* **30**, 1072-1080, (2012).

13 Michel, M. *et al.* PconsFold: improved contact predictions improve protein models. *Bioinformatics* **30**, i482-i488, (2014).

14 Ovchinnikov, S., Kamisetty, H. & Baker, D. Robust and accurate prediction of residue-residue interactions across protein interfaces using evolutionary information. *eLife* **3**, e02030, (2014).

15 Hopf, T. A. *et al.* Sequence co-evolution gives 3D contacts and structures of protein complexes. *eLife* **3**, e03430 (2014).

16 Huang, Y. J., Powers, R. & Montelione, G. T. Protein NMR recall, precision, and F-measure scores (RPF scores): structure quality assessment measures based on information retrieval statistics. *Journal of the American Chemical Society* **127**, 1665-1674, (2005).

17 Huang, Y. J., Tejero, R., Powers, R. & Montelione, G. T. A topology-constrained distance network algorithm for protein structure determination from NOESY data. *Proteins* **62**, 587-603, (2006).

18 Herrmann, T., Guntert, P. & Wuthrich, K. Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. *Journal of molecular biology* **319**, 209-227, (2002).

19 Mao, B., Tejero, R., Baker, D. & Montelione, G. T. Protein NMR structures refined with Rosetta have higher accuracy relative to corresponding X-ray crystal structures. *Journal of the American Chemical Society* **136**, 1893-1906, (2014).

20 Bhattacharya, A., Tejero, R. & Montelione, G. T. Evaluating protein structures determined by structural genomics consortia. *Proteins* **66**, 778-795, (2007).

21 Laskowski, R. A., Moss, D. S. & Thornton, J. M. Main-chain bond lengths and bond angles in protein structures. *Journal of molecular biology* **231**, 1049-1067, (1993).

22 Luthy, R., Bowie, J. U. & Eisenberg, D. Assessment of protein models with three-dimensional profiles. *Nature* **356**, 83-85, (1992).

23 Sippl, M. J. Recognition of errors in three-dimensional structures of proteins. *Proteins* **17**, 355-362, (1993).

24 Lovell, S. C. *et al.* Structure validation by Calpha geometry: phi,psi and Cbeta deviation. *Proteins* **50**, 437-450, (2003).

25 Huang, Y. J., Rosato, A., Singh, G. & Montelione, G. T. RPF: a quality assessment tool for protein NMR structures. *Nucleic Acids Res* **40**, W542-546, (2012).

26 Acton, T. B. *et al.* Preparation of protein samples for NMR structure, function, and small-molecule screening studies. *Methods Enzymol* **493**, 21-60, (2011).

27 Shen, Y. *et al.* Consistent blind protein structure generation from NMR chemical shift data. *Proc Natl Acad Sci U S A* **105**, 4685-4690, (2008).